



Sequence listing.txt
SEQUENCE LISTING

<110> Ding, Shi-You
Adney, William S.
Vinzant, Todd B.
Himmel, Michael E
Decker, Stephen R

<120> THERMAL TOLERANT CELLULASE FROM ACIDOTHERMUS CELLULOLYTICUS

<130> NREL 01-37

<140> US 09/917,383
<141> 2001-07-28

<160> 14

<170> PatentIn version 3.2

<210> 1
<211> 1228
<212> PRT
<213> Acidothermus cellulolyticus

<220>
<221> misc_feature
<223> GuxA full-length protein sequence

<400> 1

Met Glu Arg Thr Gln Gln Ser Gly Arg Asn Cys Arg Tyr Gln Arg Gly
1 5 10 15

Thr Thr Arg Met Pro Ala Ile Ser Lys Arg Leu Arg Ala Gly Val Leu
20 25 30

Ala Gly Ala Val Ser Ile Ala Ala Ser Ile Val Pro Leu Ala Met Gln
35 40 45

His Pro Ala Ile Ala Ala Thr His Val Asp Asn Pro Tyr Ala Gly Ala
50 55 60

Thr Phe Phe Val Asn Pro Tyr Trp Ala Gln Glu Val Gln Ser Glu Ala
65 70 75 80

Ala Asn Gln Thr Asn Ala Thr Leu Ala Ala Lys Met Arg Val Val Ser
85 90 95

Thr Tyr Ser Thr Ala Val Trp Met Asp Arg Ile Ala Ala Ile Asn Gly
100 105 110

Val Asn Gly Gly Pro Gly Leu Thr Thr Tyr Leu Asp Ala Ala Leu Ser
115 120 125

Gln Gln Gln Gly Thr Thr Pro Glu Val Ile Glu Ile Val Ile Tyr Asp
130 135 140

Leu Pro Gly Arg Asp Cys Ala Ala Leu Ala Ser Asn Gly Glu Leu Pro

145 150 sequence firstmg.txt 155 160

Page 2

Sequence listing.txt

420

425

430

Asp Pro Met Cys Asp Pro Thr Tyr Thr Thr Ser Tyr Gly Val Leu Thr
 435 440 445

Asn Ala Leu Pro Asn Ser Pro Ile Ala Gly Gln Trp Phe Pro Ala Gln
 450 455 460

Phe Asp Gln Leu Val Ala Asn Ala Arg Pro Ala Val Pro Thr Ser Thr
 465 470 475 480

Ser Ser Ser Pro Pro Pro Pro Pro Ser Pro Ser Ala Ser Pro Ser
 485 490 495

Pro Ser Pro Ser Pro Ser Pro Ser Ser Pro Ser Pro Ser Pro Ser
 500 505 510

Pro Ser Ser Ser Pro Ser Pro Ser Pro Ser Pro Ser Pro Ser Pro Ser
 515 520 525

Ser Ser Pro Ser Pro Ser Pro Ser Ser Ser Pro Ser Pro Ser Pro Ser
 530 535 540

Pro Ser Pro Ser Pro Ser Ser Ser Pro Ser Pro Ser Pro Ser Ser Ser
 545 550 555 560

Pro Ser Pro Ser Pro Ser Pro Ser Pro Ser Pro Ser Ser Ser Pro Ser
 565 570 575

Pro Ser Pro Thr Ser Ser Pro Val Ser Gly Gly Leu Lys Val Gln Tyr
 580 585 590

Lys Asn Asn Asp Ser Ala Pro Gly Asp Asn Gln Ile Lys Pro Gly Leu
 595 600 605

Gln Leu Val Asn Thr Gly Ser Ser Ser Val Asp Leu Ser Thr Val Thr
 610 615 620

Val Arg Tyr Trp Phe Thr Arg Asp Gly Gly Ser Ser Thr Leu Val Tyr
 625 630 635 640

Asn Cys Asp Trp Ala Ala Met Gly Cys Gly Asn Ile Arg Ala Ser Phe
 645 650 655

Gly Ser Val Asn Pro Ala Thr Pro Thr Ala Asp Thr Tyr Leu Gln Leu
 660 665 670

Ser Phe Thr Gly Gly Thr Leu Ala Ala Gly Gly Ser Thr Gly Glu Ile
 675 680 685

Gln Asn Arg Val Asn Lys Ser Asp Trp Ser Asn Phe Thr Glu Thr Asn

Sequence listing.txt

690

695

700

Asp Tyr Ser Tyr Gly Thr Asn Thr Thr Phe Gln Asp Trp Thr Lys Val
 705 710 715 720

Thr Val Tyr Val Asn Gly Val Leu Val Trp Gly Thr Glu Pro Ser Gly
 725 730 735

Thr Ser Pro Ser Pro Thr Pro Ser Pro Ser Pro Ser Pro Ser
 740 745 750

Pro Gly Gly Asp Val Thr Pro Pro Ser Val Pro Thr Gly Leu Val Val
 755 760 765

Thr Gly Val Ser Gly Ser Ser Val Ser Leu Ala Trp Asn Ala Ser Thr
 770 775 780

Asp Asn Val Gly Val Ala His Tyr Asn Val Tyr Arg Asn Gly Val Leu
 785 790 795 800

Val Gly Gln Pro Thr Val Thr Ser Phe Thr Asp Thr Gly Leu Ala Ala
 805 810 815

Gly Thr Ala Tyr Thr Tyr Thr Val Ala Ala Val Asp Ala Ala Gly Asn
 820 825 830

Thr Ser Ala Pro Ser Thr Pro Val Thr Ala Thr Thr Thr Ser Pro Ser
 835 840 845

Pro Ser Pro Thr Pro Thr Gly Thr Thr Val Thr Asp Cys Thr Pro Gly
 850 855 860

Pro Asn Gln Asn Gly Val Thr Ser Val Gln Gly Asp Glu Tyr Arg Val
 865 870 875 880

Gln Thr Asn Glu Trp Asn Ser Ser Ala Gln Gln Cys Leu Thr Ile Asn
 885 890 895

Thr Ala Thr Gly Ala Trp Thr Val Ser Thr Ala Asn Phe Ser Gly Gly
 900 905 910

Thr Gly Gly Ala Pro Ala Thr Tyr Pro Ser Ile Tyr Lys Gly Cys His
 915 920 925

Trp Gly Asn Cys Thr Thr Lys Asn Val Gly Met Pro Ile Gln Ile Ser
 930 935 940

Gln Ile Gly Ser Ala Val Thr Ser Trp Ser Thr Thr Gln Val Ser Ser
 945 950 955 960

Gly Ala Tyr Asp Val Ala Tyr Asp Ile Trp Thr Asn Ser Thr Pro Thr

Sequence listing.txt

965

970

975

Thr Thr Gly Gln Pro Asn Gly Thr Glu Ile Met Ile Trp Leu Asn Ser
 980 985 990

Arg Gly Gly Val Gln Pro Phe Gly Ser Gln Thr Ala Thr Gly Val Thr
 995 1000 1005

Val Ala Gly His Thr Trp Asn Val Trp Gln Gly Gln Gln Thr Ser
 1010 1015 1020

Trp Lys Ile Ile Ser Tyr Val Leu Thr Pro Gly Ala Thr Ser Ile
 1025 1030 1035

Ser Asn Leu Asp Leu Lys Ala Ile Phe Ala Asp Ala Ala Ala Arg
 1040 1045 1050

Gly Ser Leu Asn Thr Ser Asp Tyr Leu Leu Asp Val Glu Ala Gly
 1055 1060 1065

Phe Glu Ile Trp Gln Gly Gly Gln Gly Leu Gly Ser Asn Ser Phe
 1070 1075 1080

Ser Val Ser Val Thr Ser Gly Thr Ser Ser Pro Thr Pro Ser Pro
 1085 1090 1095

Ser Pro Thr Pro Thr Pro Ser Pro Thr Pro Thr Pro Ser Pro Ser
 1100 1105 1110

Pro Thr Pro Ser Pro Ser Pro Thr Ser Ser Pro Ser Ser Ser Gly
 1115 1120 1125

Val Ala Cys Arg Ala Thr Tyr Val Val Asn Ser Asp Trp Gly Ser
 1130 1135 1140

Gly Phe Thr Ala Thr Val Thr Val Thr Asn Thr Gly Ser Arg Ala
 1145 1150 1155

Thr Asn Gly Trp Thr Val Ala Trp Ser Phe Gly Gly Asn Gln Thr
 1160 1165 1170

Val Thr Asn Tyr Trp Asn Thr Ala Leu Thr Gln Ser Gly Ala Ser
 1175 1180 1185

Val Thr Ala Thr Asn Leu Ser Tyr Asn Asn Val Ile Gln Pro Gly
 1190 1195 1200

Gln Ser Thr Thr Phe Gly Phe Asn Gly Ser Tyr Ser Gly Thr Asn
 1205 1210 1215

Ala Ala Pro Thr Leu Ser Cys Thr Ala Ser

Sequence listing.txt

1220

1225

<210> 2
 <211> 3687
 <212> DNA
 <213> Acidothermus cellulolyticus

<220>
 <221> misc_feature
 <223> GuxA full-length coding sequence

<400> 2
 atggagcgaa cccaacaatc cggacggaac tgcaggtacc agagaggaac gacacgaatg 60
 cccgccatct caaaacggct gcgagccggc gtccctcgccg gggcgggtgag catcgagacc 120
 tccatcgtgc cgctggcgat gcagcatcct gccatcgccg cgacgcacgt cgacaatccc 180
 tatgcgggag cgaccttctt cgtcaaccgg tactgggagc aagaagtaca gagcgaagcg 240
 gcgaaccaga ccaatgccac tctcgagcg aaaaatgcgc tcgtttccac atattcgacg 300
 gccgtctgga tggaccgcat cgctgcgac aacggcgctca acggcggacc cggcttgacg 360
 acatatctgg acgccgccct ctcccagcag cagggaaacca cccctgaagt cattgagatt 420
 gtcattctacg atctgccggg acgcgactgc gcggcgctcg cctccaacgg cgaactgccc 480
 gctacggcag caggtttgca gacctatgaa acgcagtaca tcgatccgat tgcgagtatc 540
 ctgagcaatc cgaagtactc cagcctgagg atcgtgacga tcattgagcc ggactcgctg 600
 ccaaacgcgg tcaccaatat gagcattcaa gcgtgtgcaa cggcgggtgcc gtattacgag 660
 caaggcatcg agtacgcgct cacgaaattg cagccattc cgaacgtgta catctacatg 720
 gacgccgccc actccggctg gcttgggtgg cccaataatg ccagcggata cgtacaggaa 780
 gtccagaagg tcctcaacgc gagcatcggg gtcaacggca tcgacggctt cgtcaccaac 840
 acggcgaatt acacgccgtt gaaggagccg ttcattgacc ccacccagca ggtcggcgga 900
 cagccgggtg agtcggcgaa tttctaccag tggaatcctg acatcgacga agccgactac 960
 gcggttgact tgtactcgcg gctcgtcgcc gctggcttcc caagcagcat cggcatgctc 1020
 atcgacacct tacgcaacgg ttgggggtgg ccgaacgaac caacaggccc gagcaccgcg 1080
 accgatgtca acaccttcgt caaccagtcg aagattgacc ttcggcagca ccgcggcctg 1140
 tgggtgaacc agaacggtgc gggcctcggc cagccgccgc aggcaagccc gacggacttc 1200
 ccgaacgcgc acctcgacgc gtatgtctgg atcaagccgc cgggtgagtc ggacggcaca 1260
 agcgtcgcga gcgatccgac aactggcaag aagtcggacc ccatgtgcga cccgacgtac 1320
 acgacgtcgt acgggggtact gaccaacgcg ttaccgaact ccccgatcgc cggccagtgg 1380
 ttcccggcgc agtttgacca gcttgtcgcg aacgcacggc cagcggtgcc gacgtcgacc 1440
 agctcgagcc cgccgcctcc gccgccgagt ccgtcggctt cgccgagtc gagcccgagt 1500
 ccgagcccga gcagctcgcc atcgccgctg ccgtctccga gctcgagccc gtctccgtcg 1560
 ccgagcccga gtccgagccc gagtagctcg ccgtcgccgt ctccgagctc gagcccgtct 1620

Sequence listing.txt

ccgtcgccga gcccgagtcc gagcccgagt agctcgccgt cgccgtctcc gagctcgagc	1680
ccgtctccgt cgccgagccc gagtccgagc ccgagtagct cgccgtcgcc gtctccgacg	1740
tcgtcgccgg tgctgggtgg gctgaagggt cagtacaaga acaatgattc ggcgccgggt	1800
gataaccaga tcaaaccggg tctccagttg gtgaataaccg ggtcgtcgtc ggtggatttg	1860
tcgacgggtga cgggtcggtg ctggttcacc cgggatgggt ggtcgtcgac actggtgtac	1920
aactgtgact gggcgccgat ggggtgtggg aatatccgcg cctcgttcgg ctcggtgaac	1980
ccggcgacgc cgacggcgga cacctacctg cagttgtcgt tcaactgggtg aacgttggcc	2040
gctgggtgggt cgacgggtga gattcaaaac cgggtgaata agagtgactg gtcgaatttc	2100
accgagacca atgactactc gtatgggacg aacaccacct tccaggactg gacgaagggtg	2160
acgggtgtacg tcaacggcgt gttggtgtgg gggactgaac cgtccggcac cagccccagc	2220
cccacaccat ccccgagccc gagcccgagc ccgagcccgg gtggggatgt gacgccgccg	2280
agtgtgccga ccggcttggt ggtgacgggg gtgagtgggt cgtcgggtgc gttggcgtgg	2340
aatgcgtcga cggataacgt ggggggtggc cattacaacg tgtaccgcaa cggggtgttg	2400
gtgggccagc cgacgggtgac ctcggttcacc gacacgggtt tggccgcggg aaccgcgtac	2460
acctacacgg tggccgcggt ggacgctgcg ggtaacacct ccgccccatc ccccccgtc	2520
accgccacca ccacgagtcc cagccccagc cccacgccga cggggaccac ggtcaccgac	2580
tgcacgcccg gtcctaacca gaatggtgtg accagcgtgc agggcgacga ataccgggtg	2640
cagaccaatg agtggaattc gtcggcccag cagtgcctca ccatcaatac cgcgaccggt	2700
gcctggacgg tgagcactgc gaacttcagc ggtgggaccg gcggtgcgcc cgcgacgtat	2760
ccgtcgatct acaagggctg ccactggggc aactgcacca cgaagaacgt cgggatgccg	2820
attcagatca gtcagattgg ttcggctgtg acgtcgtgga gtacgacgca ggtgtcgtcg	2880
ggcgcgtatg acgtggccta cgacatttg acgaacagta cccaacgac aaccggtcag	2940
ccaaacggta ccgaaatcat gatttggtg aattcgcgtg gtggggtgca gccgttcggg	3000
tcgcagacag cgacgggtgt gacggtcgct ggtcacacgt ggaatgtctg gcagggtcag	3060
cagacctcgt ggaagattat ttcctacgtc ctgacccccg gtgcgacgtc gatcagtaat	3120
ctggatttga aggcgatttt cgcgacgcc gcggcacgcg ggtcgtcaa cacctccgat	3180
tacctgctcg acgttgaggc cgggtttgag atctggcaag gtggtcaggg cctgggcagc	3240
aactcgttca gcgtctccgt gacgagcggc acgtccagcc cgacaccgag cccgagcccg	3300
acgccgacac cgagcccgcg gccgacaccg tctccgagcc cgacccccgtc gccgagtccg	3360
accagctcgc cgtcgtcgtc ggggtgtggc tgccgggcga cgtatgtggt gaatagtgat	3420
tggggttctg ggtttacggc gacgggtgac gtgacgaata ccgggagccg ggcgacgaac	3480
gggtggacgg tggcgtggtc gtttggtggg aatcagacgg tcacgaacta ctggaacact	3540
gcgttgaccc aatcaggtgc atcgggtgac gcgacgaacc tgagttacaa caacgtgatc	3600
caaccgggtc agtcgaccac cttcggattc aacggaagtt actcaggaac aaacgccgcg	3660

ccgacgctca gctgcacagc cagctga

<210> 3
 <211> 53
 <212> PRT
 <213> Acidothermus cellulolyticus

<220>
 <221> misc_feature
 <223> potential signal peptide

<400> 3

Met Glu Arg Thr Gln Gln Ser Gly Arg Asn Cys Arg Tyr Gln Arg Gly
 1 5 10 15

Thr Thr Arg Met Pro Ala Ile Ser Lys Arg Leu Arg Ala Gly Val Leu
 20 25 30

Ala Gly Ala Val Ser Ile Ala Ala Ser Ile Val Pro Leu Ala Met Gln
 35 40 45

His Pro Ala Ile Ala
 50

<210> 4
 <211> 423
 <212> PRT
 <213> Acidothermus cellulolyticus

<220>
 <221> misc_feature
 <223> Sequence of GH6 domain

<400> 4

Ala Thr His Val Asp Asn Pro Tyr Ala Gly Ala Thr Phe Phe Val Asn
 1 5 10 15

Pro Tyr Trp Ala Gln Glu Val Gln Ser Glu Ala Ala Asn Gln Thr Asn
 20 25 30

Ala Thr Leu Ala Ala Lys Met Arg Val Val Ser Thr Tyr Ser Thr Ala
 35 40 45

Val Trp Met Asp Arg Ile Ala Ala Ile Asn Gly Val Asn Gly Gly Pro
 50 55 60

Gly Leu Thr Thr Tyr Leu Asp Ala Ala Leu Ser Gln Gln Gln Gly Thr
 65 70 75 80

Thr Pro Glu Val Ile Glu Ile Val Ile Tyr Asp Leu Pro Gly Arg Asp
 85 90 95

Cys Ala Ala Leu Ala Ser Asn Gly Glu Leu Pro Ala Thr Ala Ala Gly
 Page 8

Sequence listing.txt

100

105

110

Leu Gln Thr Tyr Glu Thr Gln Tyr Ile Asp Pro Ile Ala Ser Ile Leu
 115 120 125

Ser Asn Pro Lys Tyr Ser Ser Leu Arg Ile Val Thr Ile Ile Glu Pro
 130 135 140

Asp Ser Leu Pro Asn Ala Val Thr Asn Met Ser Ile Gln Ala Cys Ala
 145 150 155 160

Thr Ala Val Pro Tyr Tyr Glu Gln Gly Ile Glu Tyr Ala Leu Thr Lys
 165 170 175

Leu His Ala Ile Pro Asn Val Tyr Ile Tyr Met Asp Ala Ala His Ser
 180 185 190

Gly Trp Leu Gly Trp Pro Asn Asn Ala Ser Gly Tyr Val Gln Glu Val
 195 200 205

Gln Lys Val Leu Asn Ala Ser Ile Gly Val Asn Gly Ile Asp Gly Phe
 210 215 220

Val Thr Asn Thr Ala Asn Tyr Thr Pro Leu Lys Glu Pro Phe Met Thr
 225 230 235 240

Ala Thr Gln Gln Val Gly Gly Gln Pro Val Glu Ser Ala Asn Phe Tyr
 245 250 255

Gln Trp Asn Pro Asp Ile Asp Glu Ala Asp Tyr Ala Val Asp Leu Tyr
 260 265 270

Ser Arg Leu Val Ala Ala Gly Phe Pro Ser Ser Ile Gly Met Leu Ile
 275 280 285

Asp Thr Leu Arg Asn Gly Trp Gly Gly Pro Asn Glu Pro Thr Gly Pro
 290 295 300

Ser Thr Ala Thr Asp Val Asn Thr Phe Val Asn Gln Ser Lys Ile Asp
 305 310 315 320

Leu Arg Gln His Arg Gly Leu Trp Cys Asn Gln Asn Gly Ala Gly Leu
 325 330 335

Gly Gln Pro Pro Gln Ala Ser Pro Thr Asp Phe Pro Asn Ala His Leu
 340 345 350

Asp Ala Tyr Val Trp Ile Lys Pro Pro Gly Glu Ser Asp Gly Thr Ser
 355 360 365

Ala Ala Ser Asp Pro Thr Thr Gly Lys Lys Ser Asp Pro Met Cys Asp

Sequence listing.txt

370

375

380

Pro Thr Tyr Thr Thr Ser Tyr Gly Val Leu Thr Asn Ala Leu Pro Asn
385 390 395 400

Ser Pro Ile Ala Gly Gln Trp Phe Pro Ala Gln Phe Asp Gln Leu Val
405 410 415

Ala Asn Ala Arg Pro Ala Val
420

<210> 5
<211> 150
<212> PRT
<213> Acidothermus cellulolyticus

<220>
<221> misc_feature
<223> Sequence of CBD III domain

<400> 5

Val Ser Gly Gly Leu Lys Val Gln Tyr Lys Asn Asn Asp Ser Ala Pro
1 5 10 15

Gly Asp Asn Gln Ile Lys Pro Gly Leu Gln Leu Val Asn Thr Gly Ser
20 25 30

Ser Ser Val Asp Leu Ser Thr Val Thr Val Arg Tyr Trp Phe Thr Arg
35 40 45

Asp Gly Gly Ser Ser Thr Leu Val Tyr Asn Cys Asp Trp Ala Ala Met
50 55 60

Gly Cys Gly Asn Ile Arg Ala Ser Phe Gly Ser Val Asn Pro Ala Thr
65 70 75 80

Pro Thr Ala Asp Thr Tyr Leu Gln Leu Ser Phe Thr Gly Gly Thr Leu
85 90 95

Ala Ala Gly Gly Ser Thr Gly Glu Ile Gln Asn Arg Val Asn Lys Ser
100 105 110

Asp Trp Ser Asn Phe Thr Glu Thr Asn Asp Tyr Ser Tyr Gly Thr Asn
115 120 125

Thr Thr Phe Gln Asp Trp Thr Lys Val Thr Val Tyr Val Asn Gly Val
130 135 140

Leu Val Trp Gly Thr Glu
145 150

<210> 6

Sequence listing.txt

<211> 85
 <212> PRT
 <213> Acidothermus cellulolyticus

<220>
 <221> misc_feature
 <223> FN III domain of GuxA

<400> 6

Asp Val Thr Pro Pro Ser Val Pro Thr Gly Leu Val Val Thr Gly Val
 1 5 10 15

Ser Gly Ser Ser Val Ser Leu Ala Trp Asn Ala Ser Thr Asp Asn Val
 20 25 30

Gly Val Ala His Tyr Asn Val Tyr Arg Asn Gly Val Leu Val Gly Gln
 35 40 45

Pro Thr Val Thr Ser Phe Thr Asp Thr Gly Leu Ala Ala Gly Thr Ala
 50 55 60

Tyr Thr Tyr Thr Val Ala Ala Val Asp Ala Ala Gly Asn Thr Ser Ala
 65 70 75 80

Pro Ser Thr Pro Val
 85

<210> 7
 <211> 231
 <212> PRT
 <213> Acidothermus cellulolyticus

<220>
 <221> misc_feature
 <223> GH12 domain of GuxA

<400> 7

Asp Cys Thr Pro Gly Pro Asn Gln Asn Gly Val Thr Ser Val Gln Gly
 1 5 10 15

Asp Glu Tyr Arg Val Gln Thr Asn Glu Trp Asn Ser Ser Ala Gln Gln
 20 25 30

Cys Leu Thr Ile Asn Thr Ala Thr Gly Ala Trp Thr Val Ser Thr Ala
 35 40 45

Asn Phe Ser Gly Gly Thr Gly Gly Ala Pro Ala Thr Tyr Pro Ser Ile
 50 55 60

Tyr Lys Gly Cys His Trp Gly Asn Cys Thr Thr Lys Asn Val Gly Met
 65 70 75 80

Pro Ile Gln Ile Ser Gln Ile Gly Ser Ala Val Thr Ser Trp Ser Thr
 Page 11

Sequence listing.txt

85

90

95

Thr Gln Val Ser Ser Gly Ala Tyr Asp Val Ala Tyr Asp Ile Trp Thr
100 105 110

Asn Ser Thr Pro Thr Thr Thr Gly Gln Pro Asn Gly Thr Glu Ile Met
115 120 125

Ile Trp Leu Asn Ser Arg Gly Gly Val Gln Pro Phe Gly Ser Gln Thr
130 135 140

Ala Thr Gly Val Thr Val Ala Gly His Thr Trp Asn Val Trp Gln Gly
145 150 155 160

Gln Gln Thr Ser Trp Lys Ile Ile Ser Tyr Val Leu Thr Pro Gly Ala
165 170 175

Thr Ser Ile Ser Asn Leu Asp Leu Lys Ala Ile Phe Ala Asp Ala Ala
180 185 190

Ala Arg Gly Ser Leu Asn Thr Ser Asp Tyr Leu Leu Asp Val Glu Ala
195 200 205

Gly Phe Glu Ile Trp Gln Gly Gly Gln Gly Leu Gly Ser Asn Ser Phe
210 215 220

Ser Val Ser Val Thr Ser Gly
225 230

<210> 8
<211> 101
<212> PRT
<213> Acidothermus cellulolyticus

<220>
<221> misc_feature
<223> Sequence of CBD II domain

<400> 8

Gly Val Ala Cys Arg Ala Thr Tyr Val Val Asn Ser Asp Trp Gly Ser
1 5 10 15

Gly Phe Thr Ala Thr Val Thr Val Thr Asn Thr Gly Ser Arg Ala Thr
20 25 30

Asn Gly Trp Thr Val Ala Trp Ser Phe Gly Gly Asn Gln Thr Val Thr
35 40 45

Asn Tyr Trp Asn Thr Ala Leu Thr Gln Ser Gly Ala Ser Val Thr Ala
50 55 60

Thr Asn Leu Ser Tyr Asn Asn Val Ile Gln Pro Gly Gln Ser Thr Thr
Page 12

Sequence listing.txt

180

185

190

Gly Trp Leu Gly Trp Pro Asn Asn Ala Ser Gly Tyr Val Gln Glu Val
195 200 205

Gln Lys Val Leu Asn Ala Ser Ile Gly Val Asn Gly Ile Asp Gly Phe
210 215 220

Val Thr Asn Thr Ala Asn Tyr Thr Pro Leu Lys Glu Pro Phe Met Thr
225 230 235 240

Ala Thr Gln Gln Val Gly Gly Gln Pro Val Glu Ser Ala Asn Phe Tyr
245 250 255

Gln Trp Asn Pro Asp Ile Asp Glu Ala Asp Tyr Ala Val Asp Leu Tyr
260 265 270

Ser Arg Leu Val Ala Ala Gly Phe Pro Ser Ser Ile Gly Met Leu Ile
275 280 285

Asp Thr Leu Arg Asn Gly Trp Gly Gly Pro Asn Glu Pro Thr Gly Pro
290 295 300

Ser Thr Ala Thr Asp Val Asn Thr Phe Val Asn Gln Ser Lys Ile Asp
305 310 315 320

Leu Arg Gln His Arg Gly Leu Trp Cys Asn Gln Asn Gly Ala Gly Leu
325 330 335

Gly Gln Pro Pro Gln Ala Ser Pro Thr Asp Phe Pro Asn Ala His Leu
340 345 350

Asp Ala Tyr Val Trp Ile Lys Pro Pro Gly Glu Ser Asp Gly Thr Ser
355 360 365

Ala Ala Ser Asp Pro Thr Thr Gly Lys Lys Ser Asp Pro Met Cys Asp
370 375 380

Pro Thr Tyr Thr Thr Ser Tyr Gly Val Leu Thr Asn Ala Leu Pro Asn
385 390 395 400

Ser Pro Ile Ala Gly Gln Trp Phe Pro Ala Gln Phe Asp Gln Leu Val
405 410 415

Ala Asn Ala Arg Pro Ala Val
420

<210> 10
<211> 430
<212> PRT
<213> Cellulomonas fimi

Sequence listing.txt

<220>

<221> misc_feature

<223> CBHA domain of Cellulomonas fimi

<400> 10

Ala Pro Val His Val Asp Asn Pro Tyr Ala Gly Ala Val Gln Tyr Val
1 5 10 15

Asn Pro Thr Trp Ala Ala Ser Val Asn Ala Ala Ala Gly Arg Gln Ser
20 25 30

Ala Asp Pro Ala Leu Ala Ala Lys Met Arg Thr Val Ala Gly Gln Pro
35 40 45

Thr Ala Val Trp Met Asp Arg Ile Ser Ala Ile Thr Gly Asn Ala Asp
50 55 60

Gly Asn Gly Leu Lys Phe His Leu Asp Asn Ala Val Ala Gln Gln Lys
65 70 75 80

Ala Ala Gly Val Pro Leu Val Phe Asn Leu Val Ile Tyr Asp Leu Pro
85 90 95

Gly Arg Asp Cys Phe Ala Leu Ala Ser Asn Gly Glu Leu Pro Ala Thr
100 105 110

Asp Ala Gly Leu Ala Arg Tyr Lys Ser Glu Tyr Ile Asp Pro Ile Ala
115 120 125

Asp Leu Leu Asp Asn Pro Glu Tyr Glu Ser Ile Arg Ile Ala Ala Thr
130 135 140

Ile Glu Pro Asp Ser Leu Pro Asn Leu Thr Thr Asn Ile Ser Glu Pro
145 150 155 160

Ala Cys Gln Gln Ala Ala Pro Tyr Tyr Arg Gln Gly Val Lys Tyr Ala
165 170 175

Leu Asp Lys Leu His Ala Ile Pro Asn Val Tyr Asn Tyr Ile Asp Ile
180 185 190

Gly His Ser Gly Trp Leu Gly Trp Asp Ser Asn Ala Gly Pro Ser Ala
195 200 205

Thr Leu Phe Ala Glu Val Ala Lys Ser Thr Thr Ala Gly Phe Ala Ser
210 215 220

Ile Asp Gly Phe Val Ser Asp Val Ala Asn Thr Thr Pro Leu Glu Glu
225 230 235 240

Pro Leu Leu Ser Asp Ser Ser Leu Thr Ile Asn Asn Thr Pro Ile Arg
Page 15

255

Sequence listing.txt

35

40

45

Ala Val Trp Leu Asp Arg Ile Gly Ala Ile Glu Gly Asn Asp Ser Pro
50 55 60

Thr Thr Gly Ser Met Gly Leu Arg Asp His Leu Glu Glu Ala Val Arg
65 70 75 80

Gln Ser Gly Gly Asp Pro Leu Thr Ile Gln Val Val Ile Tyr Asn Leu
85 90 95

Pro Gly Arg Asp Cys Ala Ala Leu Ala Ser Asn Gly Glu Leu Gly Pro
100 105 110

Asp Glu Leu Asp Arg Tyr Lys Ser Glu Tyr Ile Asp Pro Ile Ala Asp
115 120 125

Ile Met Trp Asp Phe Ala Asp Tyr Glu Asn Leu Arg Ile Val Ala Ile
130 135 140

Ile Glu Ile Asp Ser Leu Pro Asn Leu Val Thr Asn Val Gly Gly Asn
145 150 155 160

Gly Gly Thr Glu Leu Cys Ala Tyr Met Lys Gln Asn Gly Gly Tyr Val
165 170 175

Asn Gly Val Gly Tyr Ala Leu Arg Lys Leu Gly Glu Ile Pro Asn Val
180 185 190

Tyr Asn Tyr Ile Asp Ala Ala His His Gly Trp Ile Gly Trp Asp Ser
195 200 205

Asn Phe Gly Pro Ser Val Asp Ile Phe Tyr Glu Ala Ala Asn Ala Ser
210 215 220

Gly Ser Thr Val Asp Tyr Val His Gly Phe Ile Ser Asn Thr Ala Asn
225 230 235 240

Tyr Ser Ala Thr Val Glu Pro Tyr Leu Asp Val Asn Gly Thr Val Asn
245 250 255

Gly Gln Leu Ile Arg Gln Ser Lys Trp Val Asp Trp Asn Gln Tyr Val
260 265 270

Asp Glu Leu Ser Phe Val Gln Asp Leu Arg Gln Ala Leu Ile Ala Lys
275 280 285

Gly Phe Arg Ser Asp Ile Gly Met Leu Ile Asp Thr Ser Arg Asn Gly
290 295 300

Trp Gly Gly Pro Asn Arg Pro Thr Gly Pro Ser Ser Ser Thr Asp Leu
Page 17

Sequence listing.txt

305 310 315 320

Asn Thr Tyr Val Asp Glu Ser Arg Ile Asp Arg Arg Ile His Pro Gly
325 330 335

Asn Trp Cys Asn Gln Ala Gly Ala Gly Leu Gly Glu Arg Pro Thr Val
340 345 350

Asn Pro Ala Pro Gly Val Asp Ala Tyr Val Trp Val Lys Pro Pro Gly
355 360 365

Glu Ser Asp Gly Ala Ser Glu Glu Ile Pro Asn Asp Glu Gly Lys Gly
370 375 380

Phe Asp Arg Met Cys Asp Pro Thr Tyr Gln Gly Asn Ala Arg Asn Gly
385 390 395 400

Asn Asn Pro Ser Gly Ala Leu Pro Asn Ala Pro Ile Ser Gly His Trp
405 410 415

Phe Ser Ala Gln Phe Arg Glu Leu Leu Ala Asn Ala Tyr Pro Pro Leu
420 425 430

<210> 12
<211> 221
<212> PRT
<213> Streptomyces sp.

<220>
<221> misc_feature
<223> Streptomyces sp. 11AG8 cellulase 12A

<400> 12

Asn Gln Gln Ile Cys Asp Arg Tyr Gly Thr Thr Thr Ile Gln Asp Arg
1 5 10 15

Tyr Val Val Gln Asn Asn Arg Trp Gly Thr Ser Ala Thr Gln Cys Ile
20 25 30

Asn Val Thr Gly Asn Gly Phe Glu Ile Thr Gln Ala Asp Gly Ser Val
35 40 45

Pro Thr Asn Gly Ala Pro Lys Ser Tyr Pro Ser Val Tyr Asp Gly Cys
50 55 60

His Tyr Gly Asn Cys Ala Pro Arg Thr Thr Leu Pro Met Arg Ile Ser
65 70 75 80

Ser Ile Gly Ser Ala Pro Ser Ser Val Ser Tyr Arg Tyr Thr Gly Asn
85 90 95

Gly Val Tyr Asn Ala Ala Tyr Asp Ile Trp Leu Asp Pro Thr Pro Arg
Page 18

Sequence listing.txt

100

105

110

Thr Asn Gly Val Asn Arg Thr Glu Ile Met Ile Trp Phe Asn Arg Val
115 120 125

Gly Pro Val Gln Pro Ile Gly Ser Pro Val Gly Thr Ala His Val Gly
130 135 140

Gly Arg Ser Trp Glu Val Trp Thr Gly Ser Asn Gly Ser Asn Asp Val
145 150 155 160

Ile Ser Phe Leu Ala Pro Ser Ala Ile Ser Ser Trp Ser Phe Asp Val
165 170 175

Lys Asp Phe Val Asp Gln Ala Val Ser His Gly Leu Ala Thr Pro Asp
180 185 190

Trp Tyr Leu Thr Ser Ile Gln Ala Gly Phe Glu Pro Trp Glu Gly Gly
195 200 205

Thr Gly Leu Ala Val Asn Ser Phe Ser Ser Ala Val Asn
210 215 220

<210> 13
<211> 221
<212> PRT
<213> Streptomyces lividans

<220>
<221> misc_feature
<223> Streptomyces lividans cellulase B

<400> 13

Asp Thr Thr Ile Cys Glu Pro Phe Gly Thr Thr Thr Ile Gln Gly Arg
1 5 10 15

Tyr Val Val Gln Asn Asn Arg Trp Gly Ser Thr Ala Pro Gln Cys Val
20 25 30

Thr Ala Thr Asp Thr Gly Phe Arg Val Thr Gln Ala Asp Gly Ser Ala
35 40 45

Pro Thr Asn Gly Ala Pro Lys Ser Tyr Pro Ser Val Phe Asn Gly Cys
50 55 60

His Tyr Thr Asn Cys Ser Pro Gly Thr Asp Leu Pro Val Arg Leu Asp
65 70 75 80

Thr Val Ser Ala Ala Pro Ser Ser Ile Ser Tyr Gly Phe Val Asp Gly
85 90 95

Ala Val Tyr Asn Ala Ser Tyr Asp Ile Trp Leu Asp Pro Thr Ala Arg
Page 19

Sequence listing.txt

100

105

110

Thr Asp Gly Val Asn Gln Thr Glu Ile Met Ile Trp Phe Asn Arg Val
 115 120 125

Gly Pro Ile Gln Pro Ile Gly Ser Pro Val Gly Thr Ala Ser Val Gly
 130 135 140

Gly Arg Thr Trp Glu Val Trp Ser Gly Gly Asn Gly Ser Asn Asp Val
 145 150 155 160

Leu Ser Phe Val Ala Pro Ser Ala Ile Ser Gly Trp Ser Phe Asp Val
 165 170 175

Met Asp Phe Val Arg Ala Thr Val Ala Arg Gly Leu Ala Glu Asn Asp
 180 185 190

Trp Tyr Leu Thr Ser Val Gln Ala Gly Phe Glu Pro Trp Gln Asn Gly
 195 200 205

Ala Gly Leu Ala Val Asn Ser Phe Ser Ser Thr Val Glu
 210 215 220

<210> 14
 <211> 228
 <212> PRT
 <213> Acidothermus cellulolyticus

<220>
 <221> misc_feature
 <223> Acidothermus cellulolyticus GH12 domain

<400> 14

Cys Thr Pro Gly Pro Asn Gln Asn Gly Val Thr Ser Val Gln Gly Asp
 1 5 10 15

Glu Tyr Arg Val Gln Thr Asn Glu Trp Asn Ser Ser Ala Gln Gln Cys
 20 25 30

Leu Thr Ile Asn Thr Ala Thr Gly Ala Trp Thr Val Ser Thr Ala Asn
 35 40 45

Phe Ser Gly Gly Thr Gly Gly Ala Pro Ala Thr Tyr Pro Ser Ile Tyr
 50 55 60

Lys Gly Cys His Trp Gly Asn Cys Thr Thr Lys Asn Val Gly Met Pro
 65 70 75 80

Ile Gln Ile Ser Gln Ile Gly Ser Ala Val Thr Ser Trp Ser Thr Thr
 85 90 95

Gln Val Ser Ser Gly Ala Tyr Asp Val Ala Tyr Asp Ile Trp Thr Asn
 Page 20

Sequence listing.txt

100

105

110

Ser Thr Pro Thr Thr Thr Gly Gln Pro Asn Gly Thr Glu Ile Met Ile
115 120 125

Trp Leu Asn Ser Arg Gly Gly Val Gln Pro Phe Gly Ser Gln Thr Ala
130 135 140

Thr Gly Val Thr Val Ala Gly His Thr Trp Asn Val Trp Gln Gly Gln
145 150 155 160

Gln Thr Ser Trp Lys Ile Ile Ser Tyr Val Leu Thr Pro Gly Ala Thr
165 170 175

Ser Ile Ser Asn Leu Asp Leu Lys Ala Ile Phe Ala Asp Ala Ala Ala
180 185 190

Arg Gly Ser Leu Asn Thr Ser Asp Tyr Leu Leu Asp Val Glu Ala Gly
195 200 205

Phe Glu Ile Trp Gln Gly Gly Gln Gly Leu Gly Ser Asn Ser Phe Ser
210 215 220

Val Ser Val Thr
225

FIG. 2

Diversity of glycoside hydrolase families

Glycoside Hydrolase Families	1	2	3	5	6	7	8	9	10	11	12	13	15	16	18	24	27	28	31	32	35	36	43	44	45	47	48	49	51	54	61	62	67	74
<i>Aspergillus niger</i>	•	•	•	•	•	•			•	•	•	•	•	•			•	•	•	•	•		•	•								•	•	
<i>Trichoderma reesei</i>	•	•	•	•	•	•	•			•	•	•	•	•			•						•	•								•	•	
<i>Clostridium thermocellum</i>	•		•	•	•	•			•	•	•	•		•		•								•									•	
<i>Thermobifida fusca</i>	•			•	•	•				•																							•	
<i>Cellulomonas fimi</i>	•	•		•	•	•				•	•	•				•																		
<i>Acidothermus cellulolyticus</i>				•	•	•																					•							